CUNY SPS DATA 621 - CTG5 - HW4

Betsy Rosalen, Gabrielle Bartomeo, Jeremy O'Brien, Lidiia Tronina, Rose Koh

April 24th, 2019

Contents

Table 1: Data Dictionary

VARIABLE	DEFINITION	TYPE
TARGET_FLAG	car crash = 1, no car crash = 0	binary categorical response
TARGET_AMT	car crash cost = >0, no car crash = 0	continuous numerical response
AGE	driver's age - very young/old tend to be risky	continuous numerical predictor
BLUEBOOK	$ value of vehicle	continuous numerical predictor
CAR_AGE	age of vehicle	continuous numerical predictor
CAR_TYPE	type of car (6types)	categorical predictor
CAR_USE	usage of car (commercial/private)	binary categorical predictor
CLM_FREQ	number of claims past 5 years	discrete numerical predictor
EDUCATION	max education level (5types)	categorical predictor
HOMEKIDS	number of children at home	discrete numerical predictor
HOME_VAL	$ value of home - home owners tend to drive more responsibly	continuous numerical predictor
INCOME	$ income - rich people tend to get into fewer crashes	continuous numerical predictor
JOB	job category (8types, 1missing)- white collar jobs tend to be safer	categorical predictor
KIDSDRIV	number of driving children - teenagers likely get into crashes	discrete numerical predictor
MSTATUS	maritial status - married people drive more safely	catogerical predictor
MVR_PTS	number of traffic tickets	continuous numerical predictor
OLDCLAIM	$ total claims in the past 5 years	continuous numerical predictor
PARENT1	single parent	binary categorical predictor
RED_CAR	a red car	binary categorical predictor
REVOKED	license revoked (past 7 years) - more risky driver	binary categorical predictor
SEX	gender - woman may have less crashes than man	binary categorical predictor
TIF	time in force - number of years being customer	continuous numerical predictor
TRAVTIME	distance to work	continuous numerical predictor
URBANCITY	urban/rural	binary categorical predictor
YOJ	years on job - the longer they stay more safe	continuous numerical predictor

1 DATA EXPLORATION

In this assignment we explore, analyze and model a dataset containing 8,161 observations with 25 variables each representing a customer at an auto insurance company. Two of the 25 features are target varibles and 23 are predictors. One of the target variables, `TARGET_FLAG`, is a binary categorical variable where 1 indicates that the customer has been in a car crash and 0 indicates they have not. The other target, `TARGET_AMT`, is a continuous numerical variable representing the payout amount if the customer was in a car accident. Of the remaining 23 predictor variables, 13 are categorical and 10 are numerical.

Using this data, we will compose and evaluate several types of models with the following objectives: - Logistic classification models that aim to predict the probability that a person will crash their car - Multiple linear regression models that aim to predict the amount of money it will cost if the person does crash their car

The intended use case for these models is actuarial in nature: specifically, to calculate insurance rates commensurate with policyholders' (or policy applicants') potential risk levels, based on attributes such as income, age, distance to work, tenure as customers, etc.

[JO: REWORD SO NOT REDUNDANT WITH ABOVE] [JO: DID WE CLEAN UP THE VARIABLE TYPES PER SLACK?] In the training dataset, there are 23 predictors and 2 response variables - one is binary value that indicates whether claim was made and the other is numerical value indicating the cost of claim.

[JO: CLARIFY WHAT WE MEAN BY APPROPRIATE DISTRIBUTION - CONSISTENT BETWEEN TEST AND TRAIN][JO: SHOULD WE USE A VISUAL HERE? WHAT'S THE SECOND POINT GETTING AT?]

Table 2: (#tab:t2.1)Summary statistics

	n	min	mean	median	max	sd
TARGET_AMT	8161	0	1.504325e+03	0	107586.1	4.704027e+03
AGE	8155	16	4.479031e+01	45	81.0	8.627589e+00
YOJ	7707	0	1.049929e+01	11	23.0	4.092474e+00
INCOME	7716	0	6.189809e+04	54028	367030.0	4.757268e+04
HOME_VAL	7697	0	1.548673e+05	161160	885282.0	1.291238e+05
TRAVTIME	8161	5	3.348572e+01	33	142.0	1.590833e+01
BLUEBOOK	8161	1500	1.570990e+04	14440	69740.0	8.419734e+03
TIF	8161	1	5.351305e+00	4	25.0	4.146635e+00
OLDCLAIM	8161	0	4.037076e+03	0	57037.0	8.777139e+03
MVR_PTS	8161	0	1.695503e+00	1	13.0	2.147112e+00
CAR_AGE	7651	0	8.328715e+00	8	28.0	5.700066e+00

Table 3: (#tab:t2.2)Summary statistics

	n	min	mean	median	max	sd
TARGET_FLAG	8161	0	0.2638157	0	1	0.4407276
PARENT1*	8161	1	1.1319691	1	2	0.3384779
SEX*	8161	1	1.5360863	2	2	0.4987266
MSTATUS*	8161	1	1.4003186	1	2	0.4899929
EDUCATION*	8161	1	3.0906752	3	5	1.4448565
JOB*	8161	1	5.6871707	6	9	2.6818733
CAR_TYPE*	8161	1	3.5297145	3	6	1.9653570
CAR_USE*	8161	1	1.6288445	2	2	0.4831436
RED_CAR*	8161	1	1.2913859	1	2	0.4544287
REVOKED*	8161	1	1.1225340	1	2	0.3279216
URBANICITY*	8161	1	1.2045093	1	2	0.4033673
KIDSDRIV*	8161	1	1.1710575	1	5	0.5115341
HOMEKIDS*	8161	1	1.7212351	1	6	1.1163233
CLM_FREQ*	8161	1	1.7985541	1	6	1.1584527

The response variable shows appropriate distribution in the training data. We confirm that for the number of target flags are 0 equals the target amount 0.

1.1 Summary Statistics

[JO: IN THE ENSUING SUMMARY STAT TABLES, SHALL WE REMOVE THE SCIENTIFIC NOTATION AND ROUND TO DECIMAL FOR READABILITY?][JO: WHY ARE VARIABLES SPLIT BETWEEN T2.1 AND T2.2? DO THESE PERTAIN TO THE LOGISTIC AND LINEAR MODELS, RESPECTIVELY?]

[JO: THINK THIS WOULD BE BETTER AS A SET OF SMALL MULTIPLE HISTOGRAM TABLES][JO: MSTATUS = z_F needs to be cleaned to F]

Table 4: (#tab:t2.3)Summary statistics

	TARGET_AMT	AGE	YOJ	INCOME	HOME_VAL	TRAVTIME	BLUEBOOK
	Min. : 0	Min. :16.00	Min. : 0.0	Min. : 0	Min. : 0	Min. : 5.00	Min. : 1500
	1st Qu.: 0	1st Qu.:39.00	1st Qu.: 9.0	1st Qu.: 28097	1st Qu.: 0	1st Qu.: 22.00	1st Qu.: 928
	Median : 0	Median :45.00	Median :11.0	Median : 54028	Median :161160	Median : 33.00	Median :1444
	Mean : 1504	Mean :44.79	Mean :10.5	Mean : 61898	Mean :154867	Mean : 33.49	Mean :15710
	3rd Qu.: 1036	3rd Qu.:51.00	3rd Qu.:13.0	3rd Qu.: 85986	3rd Qu.:238724	3rd Qu.: 44.00	3rd Qu.:2085
	Max. :107586	Max. :81.00	Max. :23.0	Max. :367030	Max. :885282	Max. :142.00	Max. :69740
	NA	NA's :6	NA's :454	NA's :445	NA's :464	NA	NA

Table 5: (#tab:t2.4)Summary statistics

	TARGET_FLAG	PARENT1	SEX	MSTATUS	EDUCATION	JOB	CAR_TYPE
	Min. :0.0000	No :7084	M :3786	Yes :4894	<High School :1203	z_Blue Collar:1825	Minivan :2145
	1st Qu.:0.0000	Yes:1077	z_F:4375	z_No:3267	Bachelors :2242	Clerical :1271	Panel Truck: 6
	Median :0.0000	NA	NA	NA	Masters :1658	Professional :1117	Pickup :1389
	Mean :0.2638	NA	NA	NA	PhD : 728	Manager : 988	Sports Car : 9
	3rd Qu.:1.0000	NA	NA	NA	z_High School:2330	Lawyer : 835	Van : 750
	Max. :1.0000	NA	NA	NA	NA	Student : 712	z_SUV :2294
	NA	NA	NA	NA	NA	(Other) :1413	NA

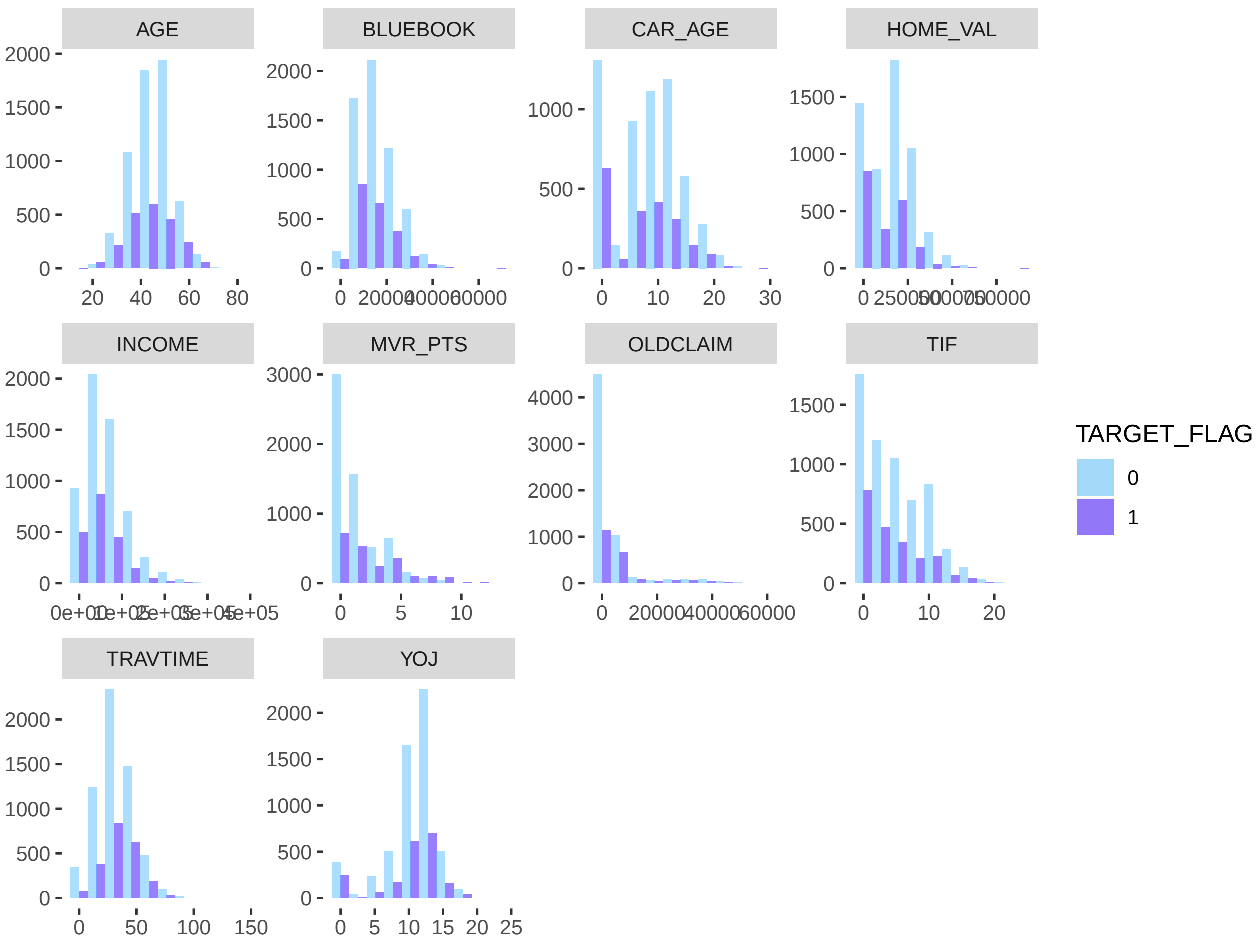


Figure 1: Numeric Data Distributions as a Function of TARGET_FLAG

[JO: IN TERMS OF HIGHER LIKELIHOODS OF ACCIDENT, LOOKS LIKE COMMERCIAL DRIVERS, BLUE COLLAR, UNMARRIED (BECAUSE YOUNGER?), PARENT, REVOKED, MALE, AND URBAN. IS IT WORTH LOOKING AT CONFUSION MATRICES?]

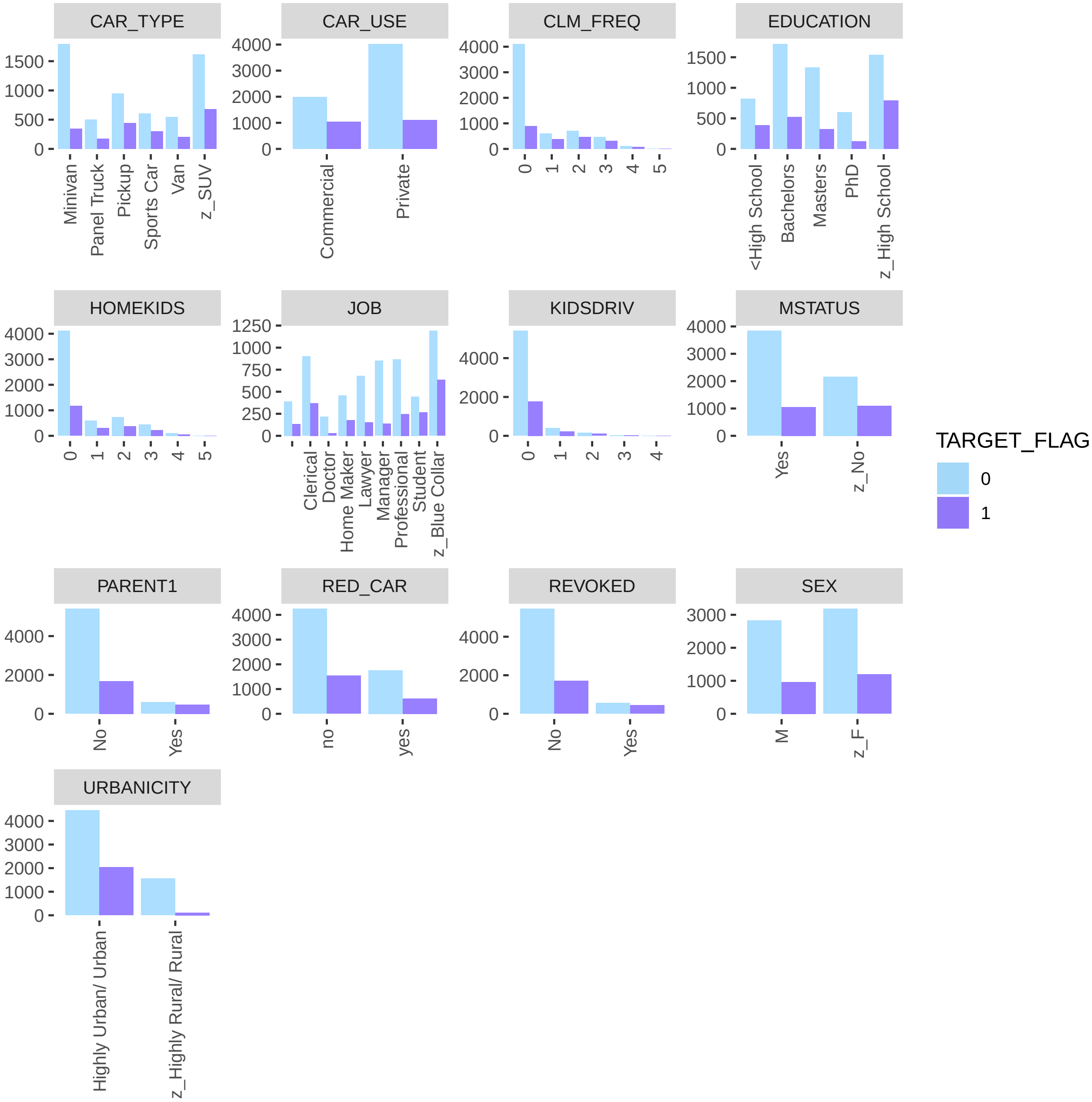


Figure 2: Categorical Data Distributions as a Function of TARGET_FLAG

[JO: IS THIS LIST BASED ONLY ON CONTINUOUS VARIABLES? PERHAPS WE SHOULD BUILD DIFFERENT GRAPHS BASED ON SCALE - ONE FOR YEARS (CAR_AGE, TIF, YOJ, AGE), DOLLARS (OLDCLAIM, BLUEBOOK, INCOME, HOMEVALUE), AND OTHER (MVR_PTS, TRAVTIME - MAYBE SEPARATE?)][JO: WHY DO BLUEBOOK AND INCOME SHOW NO DISTRO?] [JO: WHY INCLUD TARGET_AMT HERE?]

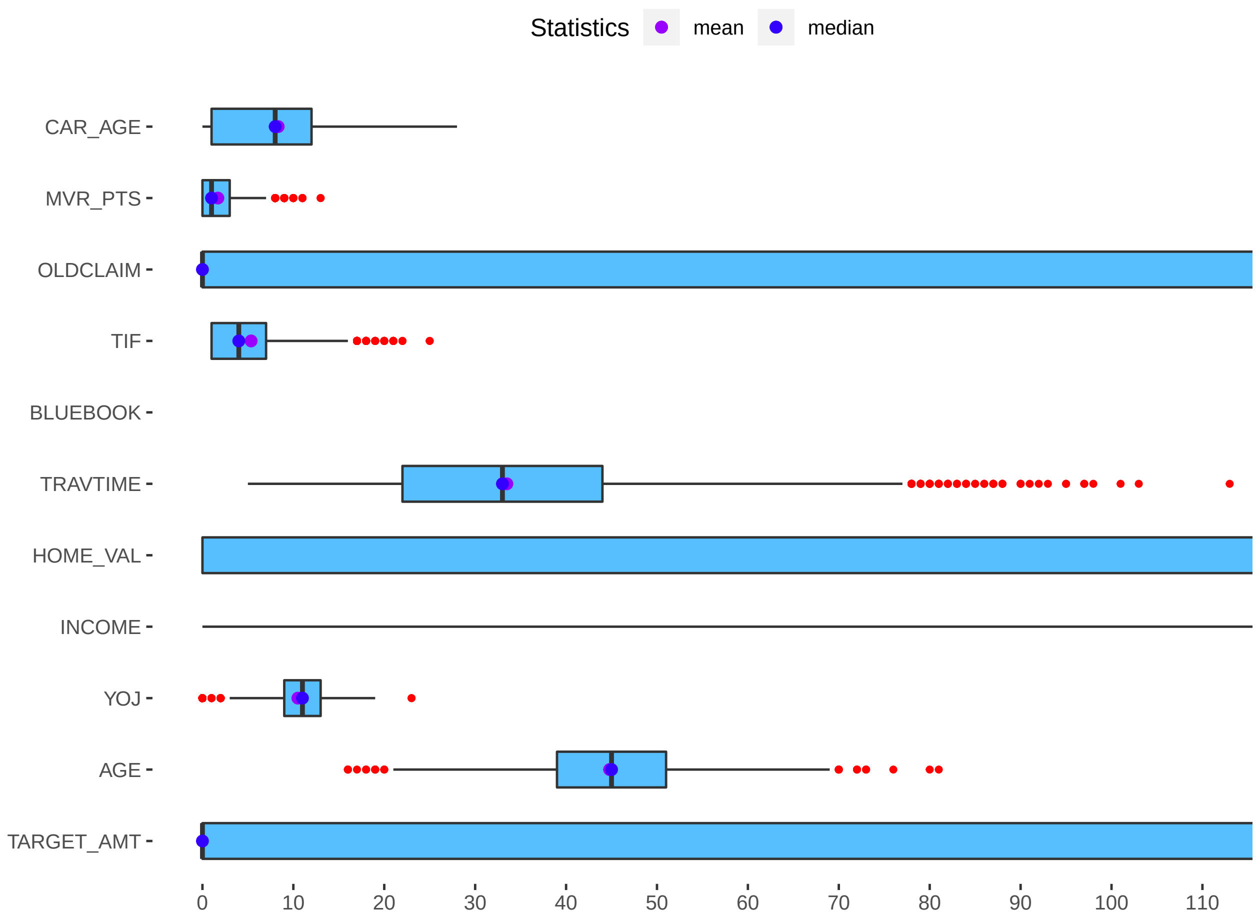


Figure 3: Outliers Boxplot

[JO: DO WE NEED THE SCALED VERSION BELOW IF WE SPLIT UP AS ABOVE? WHICH DO WE PREFER?]

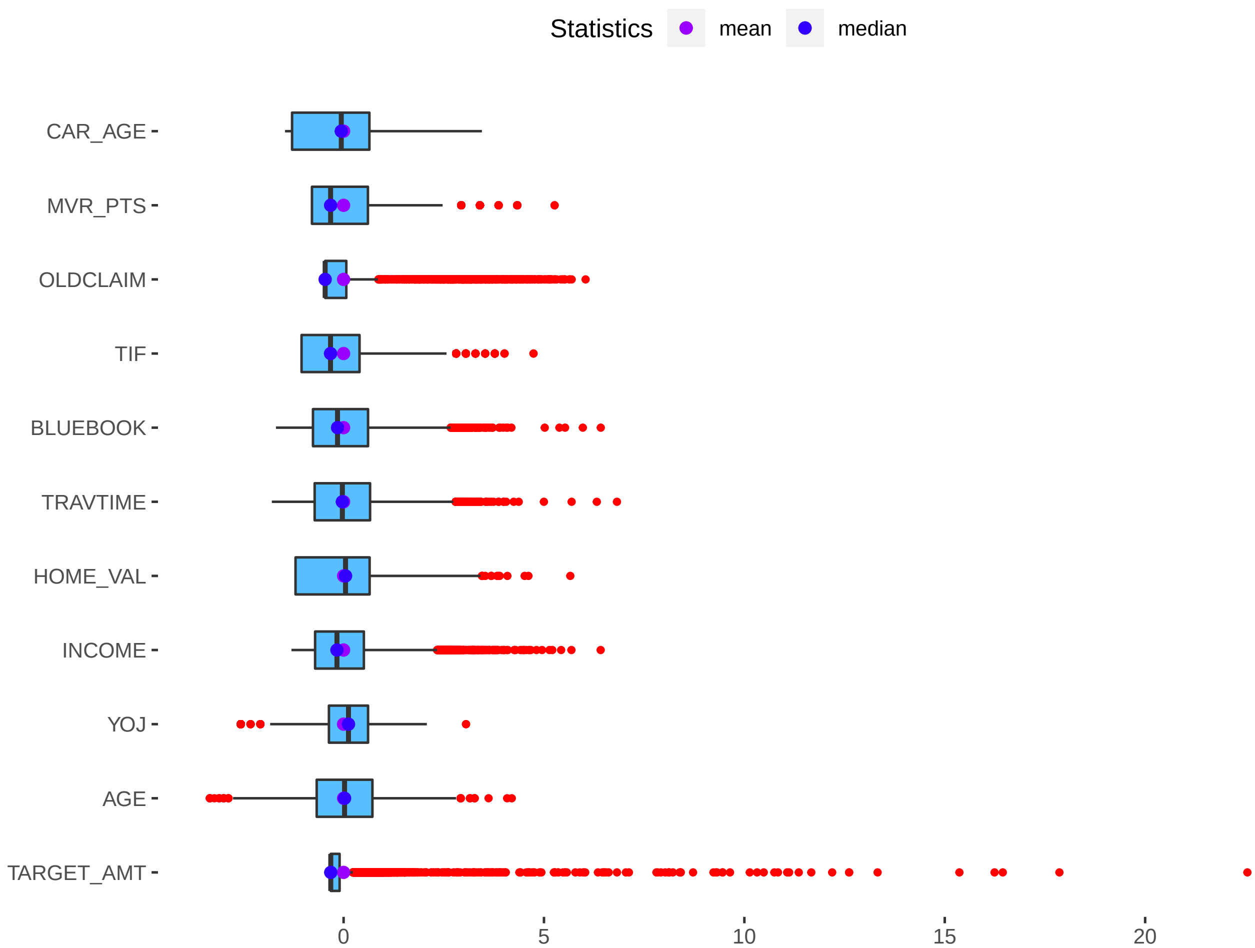


Figure 4: Scaled Boxplots

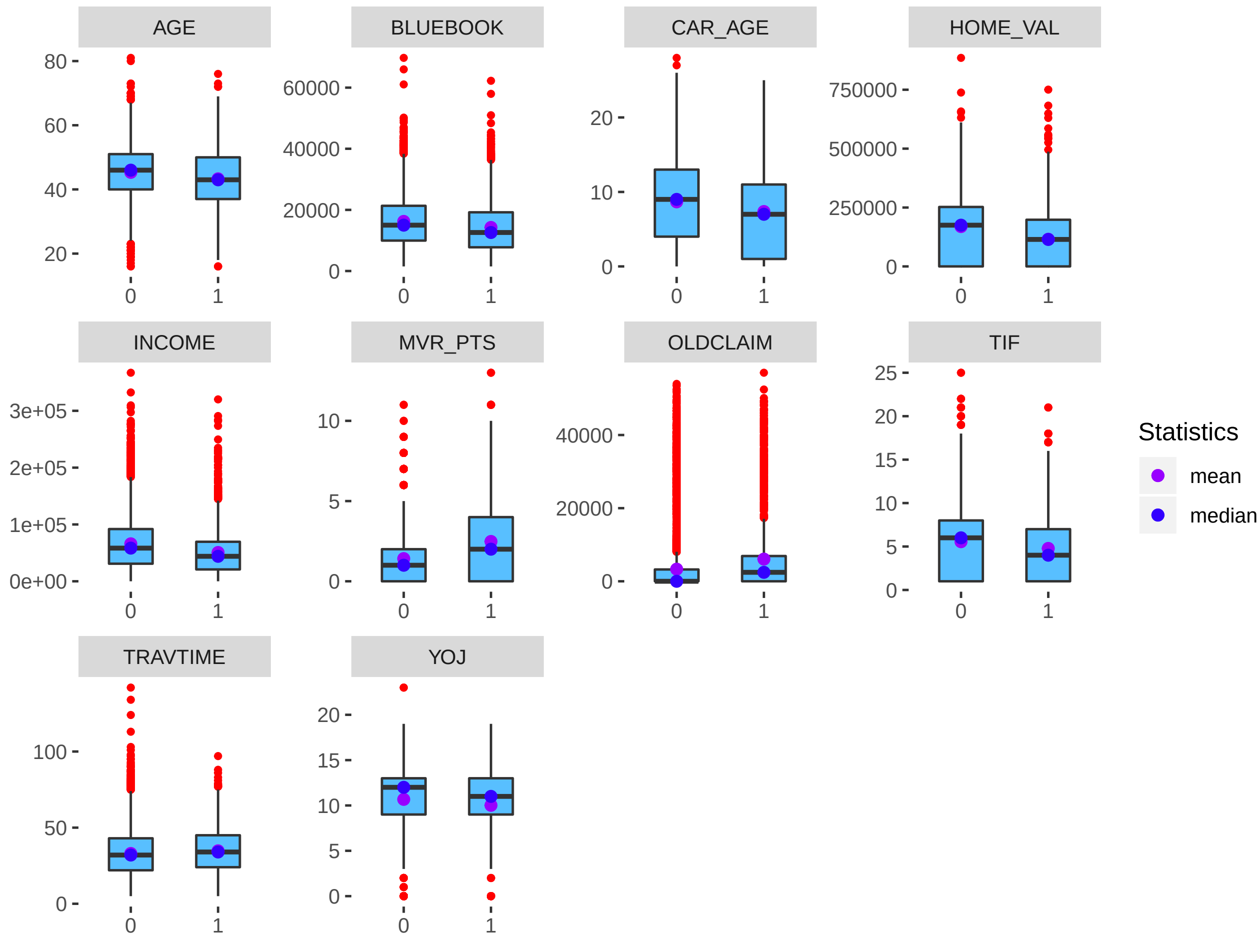


Figure 5: Linear relationship between each numeric predictor and the target

1.2 Linearity

[JO: DUE TO Y-AXIS, HARD TO TELL SLOPE - PERHAPS WE SHOULD ADD SLOPE TO THESE CHARTS TO BETTER DISCERN WHERE THERE SEEMS TO BE A LINEAR RELATIONSHIP?]

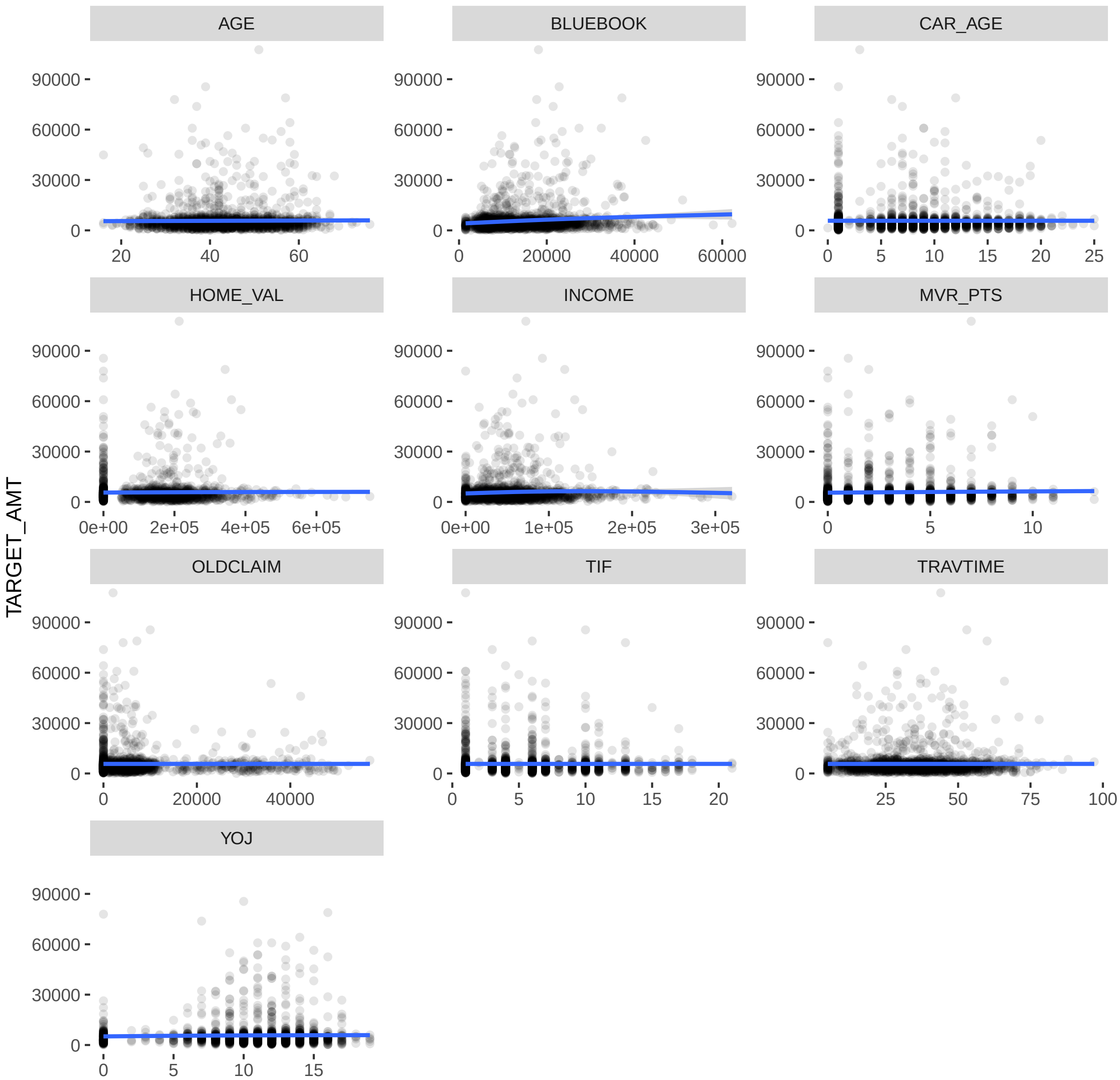


Figure 6: Scatter plot between Numeric Predictors and the TARGET_AMT

1.3 Missing Data

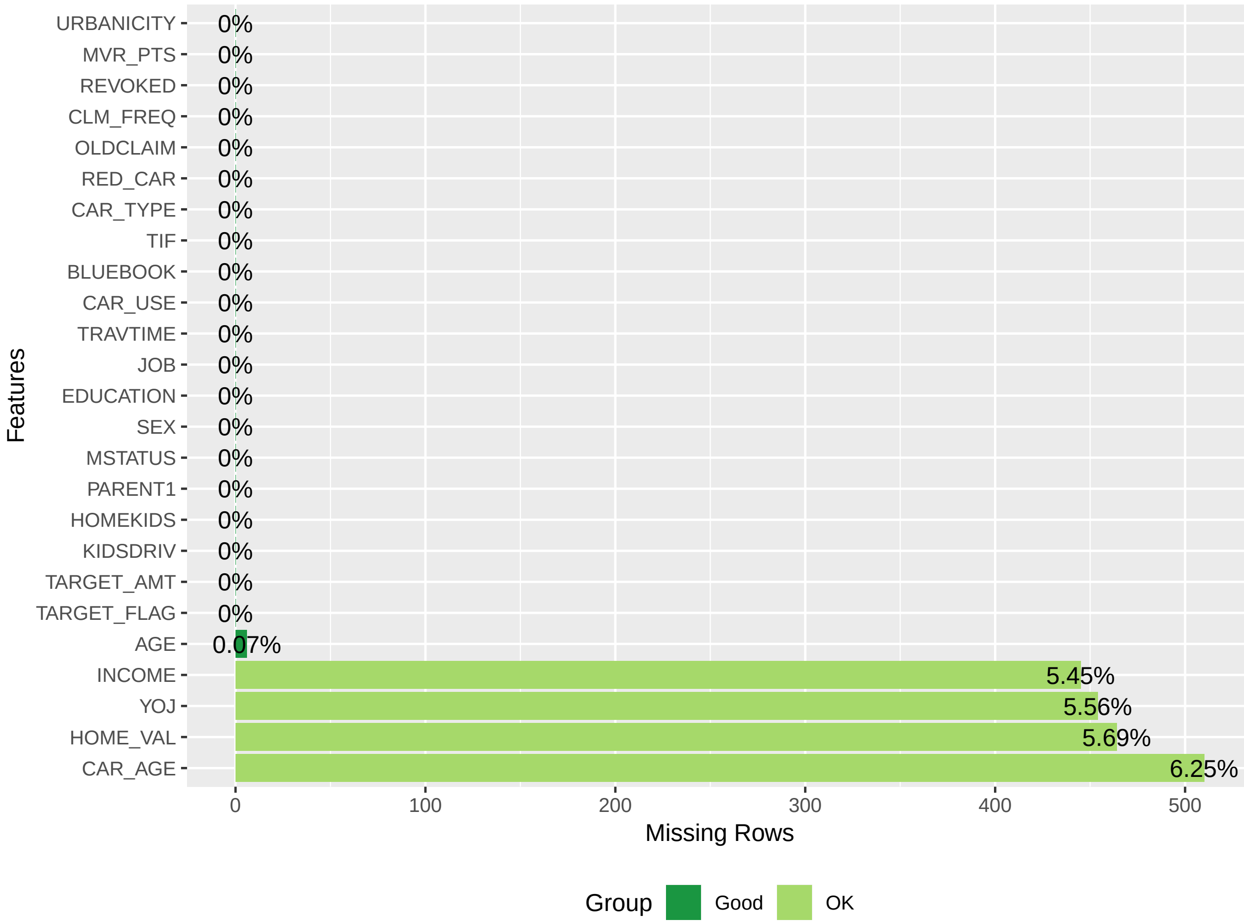


Figure 7: Missing data

There are missing observations for a number of variables: `AGE`, `INCOME`, `YOJ`, `HOME_VAL`, `CAR_AGE`.

[JO: SHALL WE INCLUDE A TABLE THAT SPELLS OUT THE MISSING PROPORTION FOR EACH VARIABLE?][JO: SUGGEST CHECKING THE OVERLAP BETWEEN MISSING VARIABLES I.E. ARE WE MISSING VARIABLES FOR THE SAME OBSERVATIONS? ADDITIONALLY, FOR THOSE MISSING VARIABLES, IS THERE SKEW OR CORRELATION IN OTHER VARIABLES I.E. THEY'RE ALL MARRIED, HIGH EARNERS, HOMEOWNERS?]

[JO: ALIGNED ON DIGGING INTO ABOVE BEFORE MAKING QUALIFICATION BELOW?] Given the low proportion, it seems acceptable to impute the missing values.

2 DATA PREPARATION

2.1 Variable Desc

2.1.0.1 KIDSDRIV

KIDSDRIV is a categorical predictor with values ranging from 0 to 4. It shows heavy skewness with most cars having 0 kid drivers. Judging from the distribution, it appears that having kid driver results in higher probability of making a claim.

2.1.0.2 AGE

AGE presents driver's age and shows normal distribution, centered around 45. Looking at the boxplot of age, there is no difference between the claim made or not in distribution. Therefore, we can believe that AGE may not be helpful in determining the probability of making a claim.

2.1.0.3 HOMEKIDS

HOMEKIDS is a predictor describing number of children at home ranging from 0 to 5.

2.1.0.4 YOJ

YOJ is a predictor describing years on job. It is believed that people who stay at a job for a long time are usually more safe. YOJ shows normal distribution apart from those who are unemployed.

2.1.0.5 INCOME

INCOME is a heavily skewed predictor variable. The outliers should be treated.

2.1.0.6 HOME_VAL

HOME_VAL is a home value predictor variable. In theory, home owners tend to drive more responsibly. In the graph, we can see difference between the owners and renters.

2.1.0.7 TRAVTIME

TRAVTIME is a predictor variable describing the distance to work. Long drives to work usually suggest greater risk. However the graph shows fairly normal distribution and it may not be helpful determining the probability of making a claim.

2.1.0.8 BLUEBOOK

BLUEBOOK is a predictor variable describing the value of the car. The‘ boxplot shows that the lower value of the car, the higher chances of making a claim. It is a possibility that the higher price cars are driven more carefully.

2.1.0.9 TIF

TIF describes how long the customer has been with the company, and the longer they have, the safer it may be. The plots show the safe drivers tend to stay safe.

2.1.0.10 OLDCLAIM

`OLDCLAIM` is a predictor describing the claims cost made in the past 5 years. We can see that it is very heavily skewed and that most people do not make claims.

2.1.0.11 CLM_FREQ

`CLM_FREQ` is a predictor that describes claim costs in the past 5 years. It seems that people who have made a claim in the past 5 years are highly likely to make another claim.

2.1.0.12 MVR_PTS

`MVR_PTS` is a predictor that describes motor vehicle record points. If you get lots of traffic tickets, you tend to get into more crash. It appears to be a highly significant variable as seen in boxplots.

2.1.0.13 CAR_AGE

`CAR_AGE` describes the vehicle age. There is one data point that shows the vehicle age is -3, this will be corrected to 0.

2.1.0.14 PARENT1

`PARENT1` describes single parent. This is factorized and renamed as `NumParents` to describe the number of parents.

2.1.0.15 SEX

`SEX` describes the gender of the driver. This is factorized and renamed as `MALE` to describe male as 1 and female as 0. It does not appear to be significant variable in the box plot.

2.1.0.16 MSTATUS

`MSTATUS` describes the martial status of the driver. It is believed that married people drive more safely. This variable has been factorized and renamed as `Single` to explain married as 0, not married as 1.

2.1.0.17 EDUCATION

`EDUCATION` describes the education level of the driver. It is factorized. It may be correlated with INCOME.

2.1.0.18 JOB

`JOB` describes the type of job the driver has. It is factorized. It may be correlated with INCOME. In theory white collar jobs tend to drive safer.

2.1.0.19 CAR_TYPE

`CAR_TYPE` describes type of car. It is factorized.

2.1.0.20 CAR_USE

`CAR_USE` describes how the car is used. Commercial vehicles are driven more and may increase probability of collision. It is factorized and renamed as `Commercial`. 0 means private.

2.1.0.21 RED_CAR

`RED_CAR` describes the color of the car is red. It is believed that red cars, especially sports cars are riskier. It is factorized.

2.1.0.22 REVOKED

`REVOKED` describes whether the license has revoked in the past 7 years. If it has revoked, it shows you are a risky driver. It is factorized. The boxplot shows the drivers who had lost their license are likely to be in accidents.

2.1.0.23 URBANICITY

`URBANICITY` describes whether driver lives in Urban area or Rural area. It is factorized and renamed as `URBAN`. 0 means rural.

[JO: THINK WE SHOULD DESCRIBE PURPOSE OF/ NEED FOR VALUE IMPUTATION. MICE IMPUTATION ASSUMES ‘MISSING AT RANDOM’ (MAR), SO THINK WE’LL NEED TO ESTABLISH THAT THIS IS THE CASE.][JO: WHAT’S THE DIFFERENCE BETWEEN THE M AND MAXIT VALUES (1 FOR AGE, 2 FOR OTHERS?)]

2.2 Missing values

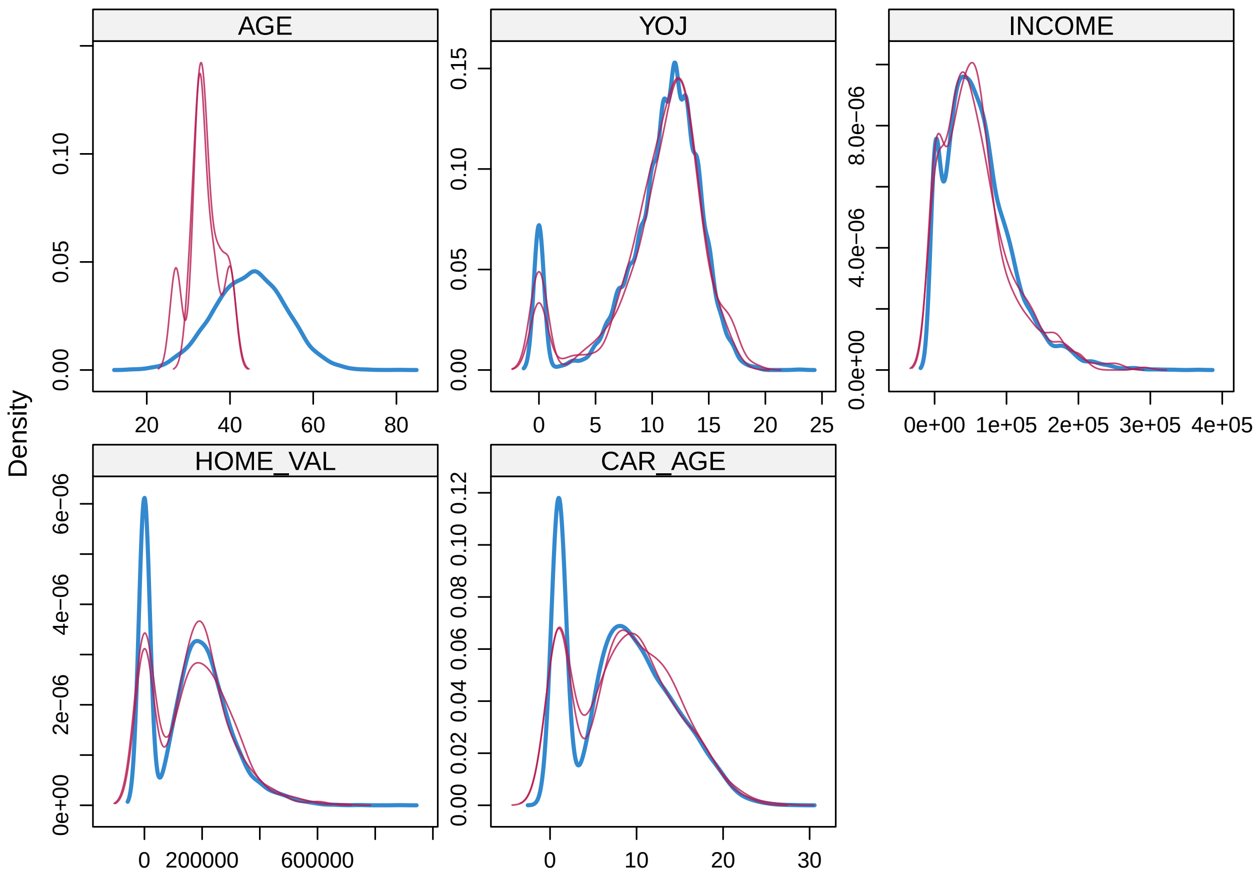
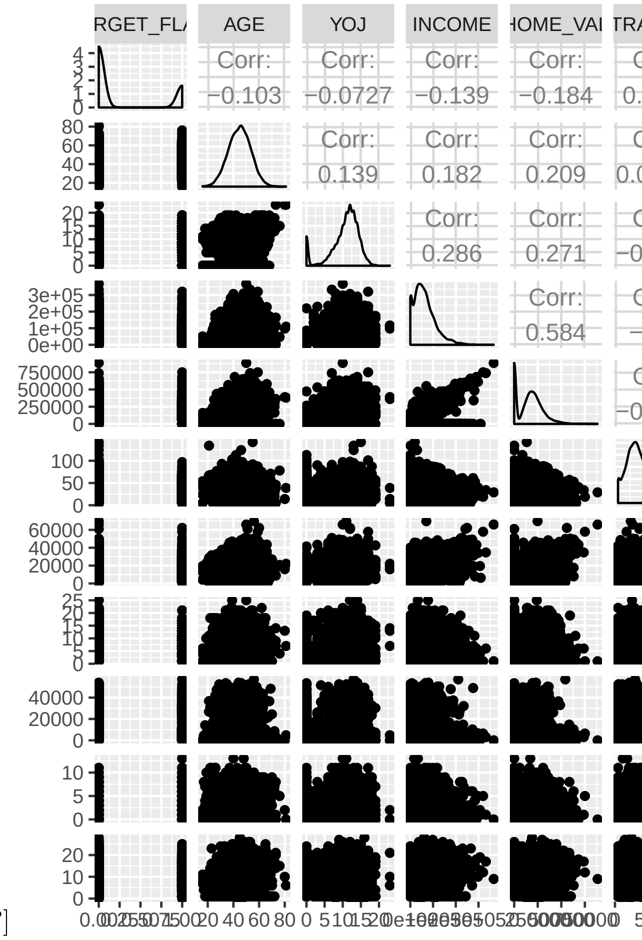


Figure 8: Difference between original and imputed data

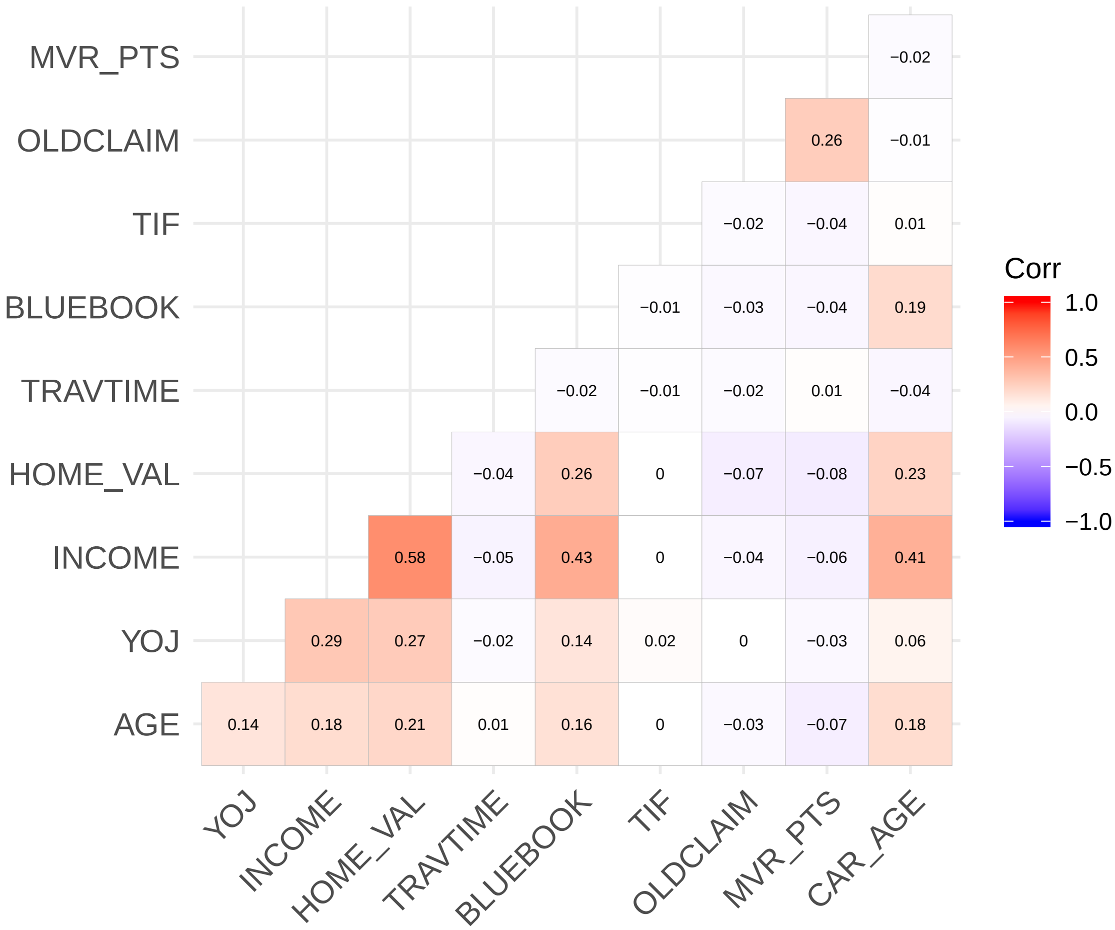
We can see that excpet the `AGE`, the 4 variables roughly matches the existing distribution. We will use the 4 variables and impute `AGE` separately, using the median imputation.

[JO: THINK THE CORR TABLE GETS MUDDLED HERE DUE TO THE LARGE NUMBER OF VARIABLES - WOULD WE CONSIDER JUST RUNNING A CORRPLOT?]



[JO: SHOULD WE THROW THIS CODE BACK IN THE SCRIPT FILE?][JO: WHY THE SUBSET OF VARIABLES HERE?]

Not surpisingly, higher levels of INCOME comes with YOJ; this also means more is disposable, which shows correlation with HOME_VAL and BLUEBOOK. [JO: HOW DOES CAR_AGE CORRELATE WITHINCOME? IS THIS DUE TO HIGHER-END VEHICLES LIVING LONGER? SEEMS SOMEWHAT COUNTERINTUITIVE] Also, MVR_PTS shows relationship with OLDCLAIMS.



MVR_PTS
OLDCLAIM
TIF
BLUEBOOK
TRAVTIME
HOME_VAL
INCOME
YOJ
AGE
YOJ
INCOME
HOME_VAL
TRAVTIME
BLUEBOOK
TIF
OLDCLAIM
MVR_PTS
CAR_AGE
Corr
1.0
0.5
0.0
−0.5
−1.0
−0.02
0.26
−0.01
−0.02
−0.04
0.01
−0.01
−0.03
−0.04
0.19
−0.02
−0.01
−0.02
0.01
−0.04
−0.04
0.26
0
−0.07
−0.08
0.23
0.58
−0.05
0.43
0
−0.04
−0.06
0.41
0.29
0.27
−0.02
0.14
0.02
0
−0.03
0.06
0.14
0.18
0.21
0.01
0.16
0
−0.03
−0.07
0.18

3 BUILD MODELS

3.1 Model 1

___ TARGET_FLAG ~ NumParents+ Male+ EDUCATION+ JOB+ CAR_TYPE+ RED_CAR+ REVOKED+ Urban+ Single+ Commercial ___

Model 1 only includes categorical variables as this will be easily interpretable and comprehensible when measuring the leading customers.

[JO: HOW SHALL WE INTERPRET THESE P-VALS - SEEM OUT OF WHACK]

```
##
## Call:
## NULL
##
## Deviance Residuals:
##       Min         1Q     Median         3Q        Max
## -0.002336   0.000000   0.000000   0.000000   0.003967
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     3.985e+02  3.339e+04   0.012    0.990
## TARGET_AMT                      1.517e+03  1.238e+04   0.123    0.902
## KIDSDRIV1                       3.310e-01  5.293e+02   0.001    1.000
## KIDSDRIV2                       2.015e-01  8.362e+02   0.000    1.000
## KIDSDRIV3                       2.806e+00  1.360e+02   0.021    0.984
## KIDSDRIV4                       1.104e+00  2.111e+03   0.001    1.000
## AGE                            -2.330e+00  3.161e+02  -0.007    0.994
## HOMEKIDS1                      -1.820e+00  1.299e+03  -0.001    0.999
## HOMEKIDS2                      -2.197e+00  5.524e+02  -0.004    0.997
## HOMEKIDS3                      -2.849e+00  4.647e+02  -0.006    0.995
## HOMEKIDS4                      -1.674e-01  6.208e+02   0.000    1.000
## HOMEKIDS5                      -3.108e+01  6.494e+03  -0.005    0.996
## YOJ                            -4.140e+00  5.194e+02  -0.008    0.994
## INCOME                         -1.598e+00  8.080e+02  -0.002    0.998
## PARENT1Yes                     -3.438e+00  9.584e+02  -0.004    0.997
## HOME_VAL                        4.646e+00  6.603e+02   0.007    0.994
## MSTATUSz_No                     3.277e+00  2.471e+02   0.013    0.989
## SEXz_F                         -2.317e+00  1.693e+03  -0.001    0.999
## EDUCATIONBachelors             -3.538e+00  8.074e+02  -0.004    0.997
## EDUCATIONMasters               -1.178e+00  2.617e+03   0.000    1.000
## EDUCATIONPhD                    2.305e+00  1.155e+03   0.002    0.998
## `EDUCATIONz_High School`        2.848e+00  3.617e+02   0.008    0.994
## JOBClerical                     6.795e+01  1.859e+04   0.004    0.997
## JOBDoctor                      -1.204e+01  1.413e+04  -0.001    0.999
## `JOBHome Maker`                 4.500e+01  1.378e+04   0.003    0.997
## JOBLawyer                       5.386e+01  1.545e+04   0.003    0.997
## JOBManager                      5.828e+01  1.675e+04   0.003    0.997
## JOBProfessional                 5.842e+01  1.774e+04   0.003    0.997
## JOBStudent                      5.081e+01  1.448e+04   0.004    0.997
## `JOBz_Blue Collar`              7.972e+01  2.136e+04   0.004    0.997
## TRAVTIME                       -1.795e+00  2.624e+02  -0.007    0.995
## CAR_USEPrivate                 -1.544e+00  3.026e+02  -0.005    0.996
## BLUEBOOK                       -1.704e+01  4.696e+02  -0.036    0.971
```

```
## TIF                                 1.117e-01  2.902e+02   0.000    1.000
## `CAR_TYPEPanel Truck`               4.313e+00  1.097e+05   0.000    1.000
## CAR_TYPEPickup                      4.005e+00  2.989e+02   0.013    0.989
## `CAR_TYPESports Car`                3.446e+00  1.126e+03   0.003    0.998
## CAR_TYPEVan                         1.931e+00  1.306e+03   0.001    0.999
## CAR_TYPEz_SUV                      -2.796e+00  1.450e+03  -0.002    0.998
## RED_CARyes                         -3.173e+00  2.450e+02  -0.013    0.990
## OLDCLAIM                           -3.020e+00  2.528e+02  -0.012    0.990
## CLM_FREQ1                           6.016e+00  3.043e+02   0.020    0.984
## CLM_FREQ2                           4.471e+00  4.570e+02   0.010    0.992
## CLM_FREQ3                           6.357e+00  3.492e+02   0.018    0.985
## CLM_FREQ4                           2.695e+00  3.751e+02   0.007    0.994
## CLM_FREQ5                           1.594e+00  3.151e+03   0.001    1.000
## REVOKEDYes                          3.305e+00  3.385e+02   0.010    0.992
## MVR_PTS                             4.043e-01  2.033e+02   0.002    0.998
## CAR_AGE                             4.084e+00  4.377e+02   0.009    0.993
## `URBANICITYz_Highly Rural/ Rural` -3.120e+00  2.894e+02  -0.011    0.991
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9.4180e+03  on 8160  degrees of freedom
## Residual deviance: 7.8951e-05  on 8111  degrees of freedom
## AIC: 100
##
## Number of Fisher Scoring iterations: 25
```

[JO: WHAT DO SINGLE TERM DELETIONS REFER TO?]

Df	Deviance	AIC
	7.54e+03	7.61e+03
1	7.54e+03	7.62e+03
1	7.54e+03	7.62e+03
1	7.62e+03	7.7e+03
4	7.58e+03	7.65e+03
8	7.65e+03	7.71e+03
5	7.67e+03	7.74e+03
1	7.61e+03	7.68e+03
1	7.54e+03	7.61e+03
1	7.62e+03	7.7e+03
1	8.11e+03	8.19e+03
4	7.58e+03	7.65e+03
5	7.55e+03	7.62e+03
5	7.67e+03	7.73e+03

[JO: CAN LIDIIA / ROSE (WHOEVER BUILT THE MODEL) ADD SOME NOTES, AND THEN I CAN ELABORATE ON THE APPROACH?]

```
##
## Call:  glm(formula = TARGET_FLAG ~ PARENT1 + SEX + MSTATUS + EDUCATION +
##     JOB + CAR_TYPE + CAR_USE + REVOKED + URBANICITY + KIDSDRIV +
##     HOMEKIDS + CLM_FREQ, family = "binomial", data = train.cat.a)
##
## Coefficients:
##                (Intercept)                        PARENT1Yes
##                   -1.67704                           0.22946
##                     SEXz_F                        MSTATUSz_No
##                   -0.29536                           0.68680
##          EDUCATIONBachelors                    EDUCATIONMasters
##                   -0.50959                          -0.44619
##               EDUCATIONPhD            EDUCATIONz_High School
##                   -0.51244                          -0.04203
##                JOBClerical                         JOBDoctor
##                    0.61911                          -0.34245
##              JOBHome Maker                         JOBLawyer
##                    0.73428                           0.18425
##                 JOBManager                    JOBProfessional
##                   -0.53115                           0.26883
##                 JOBStudent                    JOBz_Blue Collar
##                    0.76762                           0.44384
##         CAR_TYPEPanel Truck                    CAR_TYPEPickup
##                    0.17705                           0.61119
##          CAR_TYPESports Car                       CAR_TYPEVan
##                    1.23431                           0.43574
##              CAR_TYPEz_SUV                     CAR_USEPrivate
##                    0.96012                          -0.75358
##                 REVOKEDYes  URBANICITYz_Highly Rural/ Rural
##                    0.73520                          -2.22199
##                  KIDSDRIV1                         KIDSDRIV2
##                    0.46309                           0.71031
##                  KIDSDRIV3                         KIDSDRIV4
##                    1.04833                           1.41075
##                  HOMEKIDS1                         HOMEKIDS2
##                    0.33702                           0.22791
##                  HOMEKIDS3                         HOMEKIDS4
##                    0.20653                           0.04099
##                  HOMEKIDS5                         CLM_FREQ1
##                    0.39726                           0.60717
##                  CLM_FREQ2                         CLM_FREQ3
##                    0.63789                           0.65232
##                  CLM_FREQ4                         CLM_FREQ5
##                    0.90748                           0.90022
##
## Degrees of Freedom: 8160 Total (i.e. Null);  8123 Residual
## Null Deviance:       9418
## Residual Deviance: 7537  AIC: 7613
```

[JO: WHERE DOES THE RED_CAR FINDING COME IN?]

`AIC` suggests that `RED_CAR` to be removed.

	x
TARGET_AMT	1.251637e+12
KIDSDRIV1	2.286134e+09
KIDSDRIV2	5.706117e+09
KIDSDRIV3	1.508433e+08
KIDSDRIV4	3.635061e+10
AGE	8.151835e+08
HOMEKIDS1	1.377875e+10
HOMEKIDS2	2.489638e+09
HOMEKIDS3	1.762143e+09
HOMEKIDS4	3.144953e+09
HOMEKIDS5	3.441452e+11
YOJ	2.201709e+09
INCOME	5.327577e+09
PARENT1Yes	7.494753e+09
HOME_VAL	3.558199e+09
MSTATUSz_No	4.981112e+08
SEXz_F	2.339662e+10
EDUCATIONBachelors	5.319921e+09
EDUCATIONMasters	5.587443e+10
EDUCATIONPhD	1.087890e+10
‘EDUCATIONz_High School‘	1.067568e+09
JOBClerical	2.819527e+12
JOBDoctor	1.629913e+12
‘JOBHome Maker‘	1.550413e+12
JOBLawyer	1.946855e+12
JOBManager	2.289467e+12
JOBProfessional	2.567264e+12
JOBStudent	1.709784e+12
‘JOBz_Blue Collar‘	3.721698e+12
TRAVTIME	5.616486e+08
CAR_USEPrivate	7.472996e+08
BLUEBOOK	1.799763e+09
TIF	6.872163e+08
‘CAR_TYPEPanel Truck‘	9.816909e+13
CAR_TYPEPickup	7.289344e+08
‘CAR_TYPESports Car‘	1.035284e+10
CAR_TYPEVan	1.392500e+10
CAR_TYPEz_SUV	1.716795e+10
RED_CARyes	4.898158e+08
OLDCLAIM	5.215782e+08
CLM_FREQ1	7.554543e+08
CLM_FREQ2	1.703948e+09
CLM_FREQ3	9.951044e+08
CLM_FREQ4	1.147854e+09
CLM_FREQ5	8.103965e+10
REVOKEDYes	9.348191e+08
MVR_PTS	3.373373e+08
CAR_AGE	1.563454e+09
‘URBANICITYz_Highly Rural/ Rural‘	6.834244e+08

3.2 Model 2

[JO: CAN LIDIIA / ROSE (WHOEVER BUILT THE MODEL) ADD SOME NOTES, AND THEN I CAN ELABORATE ON THE APPROACH?]

```
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4441  -0.7140  -0.3890   0.6387   3.1624
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -1.443177   0.035243 -40.949  < 2e-16
## KIDSDRIV1                      0.128119   0.030758   4.165 3.11e-05
## KIDSDRIV2                      0.129712   0.029737   4.362 1.29e-05
## KIDSDRIV3                      0.076376   0.027353   2.792 0.005235
## KIDSDRIV4                      0.024204   0.026065   0.929 0.353087
## AGE                            0.018708   0.036055   0.519 0.603854
## HOMEKIDS1                      0.104755   0.037286   2.809 0.004962
## HOMEKIDS2                      0.078502   0.040343   1.946 0.051669
## HOMEKIDS3                      0.064082   0.037511   1.708 0.087568
## HOMEKIDS4                      0.016715   0.030559   0.547 0.584398
## HOMEKIDS5                      0.025316   0.027381   0.925 0.355186
## YOJ                           -0.063748   0.034708  -1.837 0.066259
## INCOME                        -0.142104   0.053557  -2.653 0.007970
## HOME_VAL                      -0.177006   0.044946  -3.938 8.21e-05
## TRAVTIME                       0.232921   0.030049   7.751 9.08e-15
## BLUEBOOK                      -0.172962   0.044471  -3.889 0.000101
## TIF                           -0.229346   0.030596  -7.496 6.58e-14
## OLDCLAIM                      -0.181395   0.037007  -4.902 9.51e-07
## CLM_FREQ1                      0.188858   0.032566   5.799 6.66e-09
## CLM_FREQ2                      0.219296   0.033171   6.611 3.82e-11
## CLM_FREQ3                      0.181649   0.031298   5.804 6.48e-09
## CLM_FREQ4                      0.122420   0.026760   4.575 4.77e-06
## CLM_FREQ5                      0.050603   0.025835   1.959 0.050148
## MVR_PTS                        0.213395   0.030245   7.056 1.72e-12
## CAR_AGE                       -0.030012   0.042837  -0.701 0.483541
## PARENT1Yes                     0.081553   0.041047   1.987 0.046945
## SEXz_F                        -0.047429   0.050127  -0.946 0.344061
## EDUCATIONBachelors            -0.163096   0.052091  -3.131 0.001742
## EDUCATIONMasters              -0.089899   0.072814  -1.235 0.216962
## EDUCATIONPhD                  -0.031657   0.061856  -0.512 0.608801
## `EDUCATIONz_High School`       0.007688   0.043027   0.179 0.858198
## JOBClerical                    0.146069   0.071728   2.036 0.041708
## JOBDoctor                     -0.076154   0.045769  -1.664 0.096136
## `JOBHome Maker`                0.057229   0.057140   1.002 0.316553
## JOBLawyer                      0.022619   0.051582   0.438 0.661026
## JOBManager                    -0.181171   0.056092  -3.230 0.001238
## JOBProfessional                0.056492   0.061545   0.918 0.358668
## JOBStudent                     0.053378   0.061122   0.873 0.382499
## `JOBz_Blue Collar`             0.135750   0.077655   1.748 0.080444
```

```
## `CAR_TYPEPanel Truck`              0.147446   0.044731   3.296 0.000980
## CAR_TYPEPickup                     0.207081   0.037952   5.456 4.86e-08
## `CAR_TYPESports Car`               0.322708   0.040919   7.887 3.11e-15
## CAR_TYPEVan                        0.175084   0.036669   4.775 1.80e-06
## CAR_TYPEz_SUV                      0.342502   0.050165   6.828 8.64e-12
## REVOKEDYes                         0.314790   0.030527  10.312  < 2e-16
## `URBANICITYz_Highly Rural/ Rural` -0.946502   0.045689 -20.716  < 2e-16
## MSTATUSz_No                        0.260945   0.043591   5.986 2.15e-09
## CAR_USEPrivate                    -0.365281   0.044466  -8.215  < 2e-16
##
## (Intercept)                       ***
## KIDSDRIV1                         ***
## KIDSDRIV2                         ***
## KIDSDRIV3                         **
## KIDSDRIV4
## AGE
## HOMEKIDS1                         **
## HOMEKIDS2                         .
## HOMEKIDS3                         .
## HOMEKIDS4
## HOMEKIDS5
## YOJ                               .
## INCOME                            **
## HOME_VAL                          ***
## TRAVTIME                          ***
## BLUEBOOK                          ***
## TIF                               ***
## OLDCLAIM                          ***
## CLM_FREQ1                         ***
## CLM_FREQ2                         ***
## CLM_FREQ3                         ***
## CLM_FREQ4                         ***
## CLM_FREQ5                         .
## MVR_PTS                           ***
## CAR_AGE
## PARENT1Yes                        *
## SEXz_F
## EDUCATIONBachelors                **
## EDUCATIONMasters
## EDUCATIONPhD
## `EDUCATIONz_High School`
## JOBClerical                       *
## JOBDoctor                         .
## `JOBHome Maker`
## JOBLawyer
## JOBManager                        **
## JOBProfessional
## JOBStudent
## `JOBz_Blue Collar`                .
## `CAR_TYPEPanel Truck`             ***
## CAR_TYPEPickup                    ***
## `CAR_TYPESports Car`              ***
## CAR_TYPEVan                       ***
## CAR_TYPEz_SUV                     ***
```

```
## REVOKEDYes                          ***
## `URBANICITYz_Highly Rural/ Rural` ***
## MSTATUSz_No                         ***
## CAR_USEPrivate                      ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9418.0  on 8160  degrees of freedom
## Residual deviance: 7262.7  on 8113  degrees of freedom
## AIC: 7358.7
##
## Number of Fisher Scoring iterations: 5
```

[JO: SAME DEAL, I CAN ELABORATE ON BASED ON SOME NOTES]

[JO: SAME DEAL, I CAN ELABORATE ON BASED ON SOME NOTES]

```
##
## Call:  glm(formula = TARGET_FLAG ~ KIDSDRIV + YOJ + INCOME + HOME_VAL +
##     TRAVTIME + BLUEBOOK + TIF + OLDCLAIM + CLM_FREQ + MVR_PTS +
##     PARENT1 + EDUCATION + JOB + CAR_TYPE + REVOKED + URBANICITY +
##     MSTATUS + CAR_USE, family = "binomial", data = train)
##
## Coefficients:
##                  (Intercept)                      KIDSDRIV1
##                   -9.822e-01                      5.883e-01
##                    KIDSDRIV2                      KIDSDRIV3
##                    7.968e-01                      9.512e-01
##                    KIDSDRIV4                            YOJ
##                    1.352e+00                     -1.308e-02
##                       INCOME                       HOME_VAL
##                   -2.982e-06                     -1.411e-06
##                     TRAVTIME                       BLUEBOOK
##                    1.455e-02                     -2.292e-05
##                          TIF                       OLDCLAIM
##                   -5.511e-02                     -2.040e-05
##                    CLM_FREQ1                      CLM_FREQ2
##                    5.766e-01                      6.242e-01
##                    CLM_FREQ3                      CLM_FREQ4
##                    6.180e-01                      8.096e-01
##                    CLM_FREQ5                        MVR_PTS
##                    1.084e+00                      1.000e-01
##                   PARENT1Yes             EDUCATIONBachelors
##                    4.391e-01                     -3.945e-01
##              EDUCATIONMasters                   EDUCATIONPhD
##                   -2.867e-01                     -1.820e-01
##        EDUCATIONz_High School                    JOBClerical
##                    1.396e-02                      4.023e-01
##                    JOBDoctor                 JOBHome Maker
##                   -4.441e-01                      1.943e-01
##                    JOBLawyer                     JOBManager
##                    6.919e-02                     -5.648e-01
##              JOBProfessional                     JOBStudent
```

```
##                     1.553e-01                         2.046e-01
##              JOBz_Blue Collar            CAR_TYPEPanel Truck
##                     3.180e-01                         5.958e-01
##                CAR_TYPEPickup             CAR_TYPESports Car
##                     5.479e-01                         9.648e-01
##                   CAR_TYPEVan                  CAR_TYPEz_SUV
##                     6.375e-01                         7.034e-01
##                    REVOKEDYes  URBANICITYz_Highly Rural/ Rural
##                     9.582e-01                        -2.348e+00
##                   MSTATUSz_No                  CAR_USEPrivate
##                     4.464e-01                        -7.509e-01
##
## Degrees of Freedom: 8160 Total (i.e. Null);  8121 Residual
## Null Deviance:       9418
## Residual Deviance: 7273  AIC: 7353
```

[JO: ON WHAT BASIS DOES THIS SUGGEST VARIABLE REMOVAL?]

`AIC` suggest to remove `AGE`, `CAR_AGE` and `Male`

	x
KIDSDRIV1	7.719982
KIDSDRIV2	7.215619
KIDSDRIV3	6.105423
KIDSDRIV4	5.543780
AGE	10.607503
HOMEKIDS1	11.344522
HOMEKIDS2	13.280597
HOMEKIDS3	11.481685
HOMEKIDS4	7.620150
HOMEKIDS5	6.117701
YOJ	9.830064
INCOME	23.405693
HOME_VAL	16.484181
TRAVTIME	7.367812
BLUEBOOK	16.137666
TIF	7.638661
OLDCLAIM	11.175538
CLM_FREQ1	8.653994
CLM_FREQ2	8.978815
CLM_FREQ3	7.993386
CLM_FREQ4	5.843561
CLM_FREQ5	5.446252
MVR_PTS	7.464283
CAR_AGE	14.973331
PARENT1Yes	13.748745
SEXz_F	20.503763
EDUCATIONBachelors	22.142077
EDUCATIONMasters	43.262830
EDUCATIONPhD	31.221788
‘EDUCATIONz_High School‘	15.106860
JOBClerical	41.982779
JOBDoctor	17.093387
‘JOBHome Maker‘	26.642116
JOBLawyer	21.711320
JOBManager	25.674234
JOBProfessional	30.908283
JOBStudent	30.485349
‘JOBz_Blue Collar‘	49.207538
‘CAR_TYPEPanel Truck‘	16.326806
CAR_TYPEPickup	11.753127
‘CAR_TYPESports Car‘	13.662746
CAR_TYPEVan	10.972163
CAR_TYPEz_SUV	20.534794
REVOKEDYes	7.604330
‘URBANICITYz_Highly Rural/ Rural‘	17.033548
MSTATUSz_No	15.505654
CAR_USEPrivate	16.134188

3.3 Model 3

```
##
```

```
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4281  -0.7144  -0.3900   0.6398   3.1777
##
## Coefficients:
##                                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         -1.44289    0.03524 -40.947  < 2e-16 ***
## KIDSDRIV1                            0.13058    0.03031   4.308 1.64e-05 ***
## KIDSDRIV2                            0.13181    0.02946   4.474 7.69e-06 ***
## KIDSDRIV3                            0.07750    0.02727   2.842 0.004477 **
## KIDSDRIV4                            0.02523    0.02584   0.976 0.328892
## HOMEKIDS1                            0.09539    0.03462   2.755 0.005871 **
## HOMEKIDS2                            0.06959    0.03778   1.842 0.065478 .
## HOMEKIDS3                            0.05547    0.03525   1.574 0.115556
## HOMEKIDS4                            0.01142    0.02940   0.389 0.697587
## HOMEKIDS5                            0.02386    0.02723   0.876 0.380972
## YOJ                                 -0.05941    0.03414  -1.740 0.081787 .
## INCOME                              -0.14614    0.05348  -2.733 0.006280 **
## HOME_VAL                            -0.17407    0.04485  -3.881 0.000104 ***
## TRAVTIME                             0.23330    0.03003   7.769 7.93e-15 ***
## BLUEBOOK                            -0.18888    0.04001  -4.721 2.34e-06 ***
## TIF                                 -0.22948    0.03059  -7.503 6.25e-14 ***
## OLDCLAIM                            -0.18139    0.03699  -4.903 9.41e-07 ***
## CLM_FREQ1                            0.18883    0.03256   5.799 6.68e-09 ***
## CLM_FREQ2                            0.21944    0.03316   6.617 3.66e-11 ***
## CLM_FREQ3                            0.18214    0.03129   5.821 5.84e-09 ***
## CLM_FREQ4                            0.12262    0.02675   4.583 4.58e-06 ***
## CLM_FREQ5                            0.05030    0.02587   1.944 0.051912 .
## MVR_PTS                              0.21281    0.03024   7.038 1.94e-12 ***
## PARENT1Yes                           0.08213    0.04102   2.002 0.045294 *
## EDUCATIONBachelors                  -0.17538    0.04886  -3.590 0.000331 ***
## EDUCATIONMasters                    -0.11139    0.06529  -1.706 0.088004 .
## EDUCATIONPhD                        -0.04631    0.05769  -0.803 0.422151
## `EDUCATIONz_High School`             0.00517    0.04287   0.121 0.904015
## JOBClerical                          0.14373    0.07168   2.005 0.044943 *
## JOBDoctor                           -0.07390    0.04572  -1.616 0.106024
## `JOBHome Maker`                      0.05318    0.05680   0.936 0.349131
## JOBLawyer                            0.02433    0.05148   0.473 0.636408
## JOBManager                          -0.18007    0.05604  -3.214 0.001311 **
## JOBProfessional                      0.05646    0.06151   0.918 0.358678
## JOBStudent                           0.05249    0.06107   0.860 0.390063
## `JOBz_Blue Collar`                   0.13486    0.07762   1.737 0.082311 .
## `CAR_TYPEPanel Truck`                0.16244    0.04179   3.887 0.000101 ***
## CAR_TYPEPickup                       0.20679    0.03791   5.454 4.92e-08 ***
## `CAR_TYPESports Car`                 0.30294    0.03386   8.946  < 2e-16 ***
## CAR_TYPEVan                          0.18412    0.03543   5.196 2.04e-07 ***
## CAR_TYPEz_SUV                        0.31338    0.03881   8.074 6.78e-16 ***
## REVOKEDYes                           0.31476    0.03051  10.318  < 2e-16 ***
## `URBANICITYz_Highly Rural/ Rural` -0.94647    0.04569 -20.715  < 2e-16 ***
## MSTATUSz_No                          0.25943    0.04349   5.965 2.45e-09 ***
## CAR_USEPrivate                      -0.36513    0.04443  -8.218  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9418.0  on 8160  degrees of freedom
## Residual deviance: 7264.5  on 8116  degrees of freedom
## AIC: 7354.5
##
## Number of Fisher Scoring iterations: 5
```

```
##
## Call:  glm(formula = TARGET_FLAG ~ KIDSDRIV + YOJ + INCOME + HOME_VAL +
##     TRAVTIME + BLUEBOOK + TIF + OLDCLAIM + CLM_FREQ + MVR_PTS +
##     PARENT1 + EDUCATION + JOB + CAR_TYPE + REVOKED + URBANICITY +
##     MSTATUS + CAR_USE, family = "binomial", data = train)
##
## Coefficients:
##                   (Intercept)                       KIDSDRIV1
##                    -9.822e-01                       5.883e-01
##                     KIDSDRIV2                       KIDSDRIV3
##                     7.968e-01                       9.512e-01
##                     KIDSDRIV4                             YOJ
##                     1.352e+00                      -1.308e-02
##                        INCOME                        HOME_VAL
##                    -2.982e-06                      -1.411e-06
##                      TRAVTIME                        BLUEBOOK
##                     1.455e-02                      -2.292e-05
##                           TIF                        OLDCLAIM
##                    -5.511e-02                      -2.040e-05
##                     CLM_FREQ1                       CLM_FREQ2
##                     5.766e-01                       6.242e-01
##                     CLM_FREQ3                       CLM_FREQ4
##                     6.180e-01                       8.096e-01
##                     CLM_FREQ5                         MVR_PTS
##                     1.084e+00                       1.000e-01
##                    PARENT1Yes             EDUCATIONBachelors
##                     4.391e-01                      -3.945e-01
##              EDUCATIONMasters                    EDUCATIONPhD
##                    -2.867e-01                      -1.820e-01
##        EDUCATIONz_High School                     JOBClerical
##                     1.396e-02                       4.023e-01
##                     JOBDoctor                   JOBHome Maker
##                    -4.441e-01                       1.943e-01
##                     JOBLawyer                      JOBManager
##                     6.919e-02                      -5.648e-01
##               JOBProfessional                      JOBStudent
##                     1.553e-01                       2.046e-01
##              JOBz_Blue Collar            CAR_TYPEPanel Truck
##                     3.180e-01                       5.958e-01
##                CAR_TYPEPickup              CAR_TYPESports Car
##                     5.479e-01                       9.648e-01
##                   CAR_TYPEVan                    CAR_TYPEz_SUV
##                     6.375e-01                       7.034e-01
```

```
##                      REVOKEDYes  URBANICITYz_Highly Rural/ Rural
##                       9.582e-01                       -2.348e+00
##                     MSTATUSz_No                   CAR_USEPrivate
##                       4.464e-01                       -7.509e-01
##
## Degrees of Freedom: 8160 Total (i.e. Null);  8121 Residual
## Null Deviance:       9418
## Residual Deviance: 7273  AIC: 7353
```

	x
KIDSDRIV1	7.495313
KIDSDRIV2	7.083295
KIDSDRIV3	6.066746
KIDSDRIV4	5.447691
HOMEKIDS1	9.782557
HOMEKIDS2	11.646107
HOMEKIDS3	10.138444
HOMEKIDS4	7.054505
HOMEKIDS5	6.050739
YOJ	9.508996
INCOME	23.335243
HOME_VAL	16.414733
TRAVTIME	7.358960
BLUEBOOK	13.060424
TIF	7.633984
OLDCLAIM	11.166587
CLM_FREQ1	8.652728
CLM_FREQ2	8.973892
CLM_FREQ3	7.988530
CLM_FREQ4	5.840954
CLM_FREQ5	5.462976
MVR_PTS	7.459405
PARENT1Yes	13.732822
EDUCATIONBachelors	19.476277
EDUCATIONMasters	34.788815
EDUCATIONPhD	27.161546
‘EDUCATIONz_High School‘	14.998438
JOBClerical	41.924381
JOBDoctor	17.056386
‘JOBHome Maker‘	26.322206
JOBLawyer	21.624062
JOBManager	25.621481
JOBProfessional	30.874901
JOBStudent	30.429519
‘JOBz_Blue Collar‘	49.163798
‘CAR_TYPEPanel Truck‘	14.248140
CAR_TYPEPickup	11.730468
‘CAR_TYPESports Car‘	9.356915
CAR_TYPEVan	10.245672
CAR_TYPEz_SUV	12.291240
REVOKEDYes	7.594472
‘URBANICITYz_Highly Rural/ Rural‘	17.034479
MSTATUSz_No	15.435664
CAR_USEPrivate	16.108458

3.4 Model 4 - Binary Logistic model

The forth model is a binary logistic model including all the explanatory variables plus log transformations of our skewed variables: income, travtime, bluebook, oldclaim and age. We used the backward elimination function to refine our model. *Lidiia: I am not sure why, but there is a formatting issue. The summary for this and all other models below are at the end of the pdf.*

3.5 Model 5 - Multiple linear regression model

A multiple linear regression model using only positive target_amt in our training and stepwise elimination.

3.6 Model 6 - Multiple linear regression model

A multiple linear regression model using all cases including $0 target_amt. We used the backward elimination function to refine our model. Any predicted value less than $100 will be considered 0.

4 SELECT MODELS

5 Appendix

The appendix is available as script.R file in `project4_insurance` folder.

https://github.com/betsyrosalen/DATA_621_Business_Analyt_and_Data_Mining

Df	Deviance	AIC
	7.26e+03	7.36e+03
4	7.3e+03	7.39e+03
1	7.26e+03	7.36e+03
5	7.27e+03	7.36e+03
1	7.27e+03	7.36e+03
1	7.27e+03	7.36e+03
1	7.28e+03	7.37e+03
1	7.32e+03	7.42e+03
1	7.28e+03	7.37e+03
1	7.32e+03	7.41e+03
1	7.29e+03	7.38e+03
5	7.33e+03	7.41e+03
1	7.31e+03	7.41e+03
1	7.26e+03	7.36e+03
1	7.27e+03	7.36e+03
1	7.26e+03	7.36e+03
4	7.28e+03	7.37e+03
8	7.32e+03	7.4e+03
5	7.35e+03	7.44e+03
1	7.37e+03	7.46e+03
1	7.87e+03	7.97e+03
1	7.3e+03	7.39e+03
1	7.33e+03	7.43e+03

Df	Deviance	AIC
	7.26e+03	7.35e+03
4	7.3e+03	7.38e+03
5	7.27e+03	7.35e+03
1	7.27e+03	7.36e+03
1	7.27e+03	7.36e+03
1	7.28e+03	7.37e+03
1	7.33e+03	7.41e+03
1	7.29e+03	7.38e+03
1	7.32e+03	7.41e+03
1	7.29e+03	7.38e+03
5	7.33e+03	7.41e+03
1	7.31e+03	7.4e+03
1	7.27e+03	7.36e+03
4	7.29e+03	7.37e+03
8	7.32e+03	7.4e+03
5	7.37e+03	7.45e+03
1	7.37e+03	7.46e+03
1	7.87e+03	7.96e+03
1	7.3e+03	7.39e+03
1	7.33e+03	7.42e+03

Observations	8161
Dependent variable	TARGET_FLAG
Type	Generalized linear model
Family	binomial
Link	logit

$\chi^2(35)$	2140.47
Pseudo-R^2 (Cragg-Uhler)	0.34
Pseudo-R^2 (McFadden)	0.23
AIC	7349.49
BIC	7601.75

	Est.	S.E.	z val.	p	VIF
(Intercept)	2.04	0.79	2.57	0.01	NA
KIDSDRIV1	0.58	0.11	5.54	0.00	1.15
KIDSDRIV2	0.82	0.15	5.40	0.00	1.15
KIDSDRIV3	0.98	0.30	3.25	0.00	1.15
KIDSDRIV4	1.38	1.12	1.23	0.22	1.15
log(AGE)	-0.31	0.16	-2.00	0.05	1.26
YOJ	0.01	0.01	1.30	0.19	2.37
log(INCOME + 1e-14)	-0.02	0.00	-4.07	0.00	3.23
HOME_VAL	-0.00	0.00	-5.25	0.00	1.75
log(TRAVTIME)	0.41	0.05	7.94	0.00	1.03
log(BLUEBOOK)	-0.32	0.06	-5.87	0.00	1.48
TIF	-0.05	0.01	-7.34	0.00	1.01
log(OLDCLAIM + 1e-14)	0.01	0.00	6.24	0.00	1.26
MVR_PTS	0.10	0.01	7.09	0.00	1.24
PARENT1Yes	0.37	0.10	3.66	0.00	1.64
EDUCATIONBachelors	-0.41	0.11	-3.75	0.00	7.47
EDUCATIONMasters	-0.33	0.16	-2.04	0.04	7.47
EDUCATIONPhD	-0.30	0.19	-1.53	0.13	7.47
EDUCATIONz_High School	0.03	0.09	0.35	0.73	7.47
JOBClerical	0.49	0.19	2.53	0.01	26.60
JOBDoctor	-0.39	0.27	-1.48	0.14	26.60
JOBHome Maker	0.15	0.21	0.69	0.49	26.60
JOBLawyer	0.16	0.17	0.96	0.34	26.60
JOBManager	-0.51	0.17	-2.99	0.00	26.60
JOBProfessional	0.22	0.18	1.26	0.21	26.60
JOBStudent	0.12	0.22	0.55	0.58	26.60
JOBz_Blue Collar	0.39	0.19	2.10	0.04	26.60
CAR_TYPEPanel Truck	0.54	0.14	3.76	0.00	2.33
CAR_TYPEPickup	0.58	0.10	5.80	0.00	2.33
CAR_TYPESports Car	0.96	0.11	8.85	0.00	2.33
CAR_TYPEVan	0.65	0.12	5.32	0.00	2.33
CAR_TYPEz_SUV	0.74	0.09	8.57	0.00	2.33
REVOKEDYes	0.71	0.08	8.82	0.00	1.01
URBANICITYz_Highly Rural/ Rural	-2.35	0.11	-20.86	0.00	1.14
MSTATUSz_No	0.46	0.08	5.62	0.00	1.96
CAR_USEPrivate	-0.75	0.09	-8.18	0.00	2.46

Standard errors: MLE

Observations	2153
Dependent variable	TARGET_FLAG
Type	OLS linear regression

F(8,2144)	268.01
R^2	0.50
Adj. R^2	0.50

	Est.	S.E.	t val.	p
(Intercept)	1.00	0.00	336544377181634.00	0.00
INCOME	-0.00	0.00	-2.82	0.00
HOME_VAL	0.00	0.00	2.12	0.03
CAR_TYPEPanel Truck	0.00	0.00	1.00	0.32
CAR_TYPEPickup	0.00	0.00	0.47	0.64
CAR_TYPESports Car	-0.00	0.00	-2.19	0.03
CAR_TYPEVan	0.00	0.00	0.69	0.49
CAR_TYPEz_SUV	-0.00	0.00	-0.22	0.83
CAR_USEPrivate	0.00	0.00	1.80	0.07

Standard errors: OLS

Observations	8161
Dependent variable	TARGET_FLAG
Type	OLS linear regression

F(36,8124)	71.15
R^2	0.24
Adj. R^2	0.24

	Est.	S.E.	t val.	p
(Intercept)	3.99	0.41	9.63	0.00
KIDSDRIV1	0.11	0.02	6.50	0.00
KIDSDRIV2	0.14	0.02	5.95	0.00
KIDSDRIV3	0.18	0.05	3.70	0.00
KIDSDRIV4	0.15	0.19	0.77	0.44
log(AGE)	-1.23	0.15	-8.33	0.00
AGE	0.03	0.00	8.15	0.00
log(INCOME + 1e-14)	-0.00	0.00	-4.23	0.00
HOME_VAL	-0.00	0.00	-4.70	0.00
log(TRAVTIME)	0.05	0.01	7.66	0.00
log(BLUEBOOK)	-0.05	0.01	-5.66	0.00
TIF	-0.01	0.00	-7.47	0.00
log(OLDCLAIM + 1e-14)	0.00	0.00	7.62	0.00
OLDCLAIM	-0.00	0.00	-5.06	0.00
MVR_PTS	0.02	0.00	7.84	0.00
PARENT1Yes	0.06	0.02	3.90	0.00
EDUCATIONBachelors	-0.06	0.02	-3.69	0.00
EDUCATIONMasters	-0.04	0.02	-1.74	0.08
EDUCATIONPhD	-0.04	0.03	-1.62	0.11
EDUCATIONz_High School	0.01	0.01	0.48	0.63
JOBClerical	0.10	0.03	3.58	0.00
JOBDoctor	-0.04	0.03	-1.18	0.24
JOBHome Maker	0.06	0.03	1.81	0.07
JOBLawyer	0.03	0.03	1.27	0.20
JOBManager	-0.06	0.02	-2.32	0.02
JOBProfessional	0.06	0.03	2.13	0.03
JOBStudent	0.05	0.03	1.40	0.16
JOBz_Blue Collar	0.09	0.03	3.28	0.00
CAR_TYPEPanel Truck	0.06	0.02	2.57	0.01
CAR_TYPEPickup	0.07	0.01	5.03	0.00
CAR_TYPESports Car	0.12	0.02	7.59	0.00
CAR_TYPEVan	0.08	0.02	4.58	0.00
CAR_TYPEz_SUV	0.09	0.01	7.77	0.00
REVOKEDYes	0.16	0.01	10.78	0.00
URBANICITYz_Highly Rural/ Rural	-0.29	0.01	-24.23	0.00
MSTATUSz_No	0.07	0.01	6.02	0.00
CAR_USEPrivate	-0.12	0.01	-8.57	0.00

Standard errors: OLS